

AMENDMENTS TO CLAIMS:

The listing of claims will replace all prior versions, and listings, of claims in the application:

LISTING OF CLAIMS:

1. (Currently Amended) A document categorizing method for categorizing a plurality of documents into a plurality of clusters according to semantic similarity, said method being characterized in that:

after categorizing said plurality of documents into a plurality of clusters according to semantic similarity, a cluster merging process is performed such that relations among clusters of said plurality of clusters are evaluated on the basis of documents included in the respective clusters, and two or more clusters having a degree of relation equal to or higher than a predetermined value are combined together;

wherein said cluster merging process defines said degree of relation between multiple clusters under consideration as the number of distinct files common to all of said clusters under consideration multiplied by a predefined multiplication factor divided by a total sum of all the files in said clusters under consideration.

2. (Currently Amended) A document categorizing method according to Claim 1, wherein said multiplication factor is equal to the number of clusters under consideration. said cluster merging process is performed such that the evaluation of relations among clusters under consideration as to whether they should be merged or not is performed on the basis of the number of documents commonly included in said clusters under consideration relative to the total number of documents included in said clusters under consideration, and cluster merging is performed in accordance with the evaluation result.

3. (Previously Presented) A document categorizing method according to Claim 1, wherein said cluster merging process is performed such that the manner in which feature elements, which characterize respective clusters under consideration as to whether they should be merged or not, appear in said

respective clusters under consideration is examined, and cluster merging is performed in accordance with the manner in which the feature elements appear.

4. (Previously Presented) A document categorizing method according to Claim 1, wherein said cluster merging process is performed on at least two clusters, and after completion of said cluster merging process a first time, said cluster merging process is repeatedly performed on the resultant set of clusters until no further cluster merging occurs.

5. (Previously Presented) A document categorizing method according to Claim 1, wherein after completion of said cluster merging process, supplementary information indicating that cluster merging has been performed and also indicating the basis on which the cluster merging has been performed is output.

6. (Cancelled)

7. (Currently Amended) A document categorizing method ~~according to Claim 6, for categorizing a plurality of documents into a plurality of clusters according to semantic similarity, said method being characterized in that:~~

~~after categorizing said plurality of documents into a plurality of clusters according to semantic similarity, a cluster merging process is performed such that relations among clusters of said plurality of clusters are evaluated on the basis of documents included in the respective clusters, and two or more clusters having a degree of relation equal to or higher than a first predetermined value are combined together;~~

~~wherein said generated information, which is output so as to enable the user to see in what manner said cluster merging process has been performed, is presented by modifying the manner of displaying the cluster names of respective clusters merged together in accordance with the degree of relation among said clusters merged together, said cluster names being are displayed in such a manner that when said degree of relation among said clusters is higher than a second predetermined value higher than said first predetermined value, said cluster names are displayed in an a AND first listing format, and when said degree of relation among said clusters is lower than said second predetermined~~

value and higher than said first predetermined value, said cluster names are displayed in an OR-a second listing format.

8. (Currently Amended) A document categorizing method according to Claim 7, wherein when said cluster names are displayed in said AND-first listing format, said cluster names of the respective clusters are displayed successively in a single horizontal line or are displayed successively in different lines, and when said cluster names are displayed in said OR-second listing format, a delimiter is inserted between adjacent cluster names of the respective clusters.

9. (Previously Presented) A document categorizing method according to Claim 7, wherein when a first cluster includes a second cluster therein, the name of said second cluster included in said first cluster is enclosed within brackets and placed after the name of said first cluster.

10. (Currently Amended) A document categorizing apparatus for categorizing a plurality of documents into a plurality of clusters according to semantic similarity, said apparatus comprising:

a clustering unit for categorizing a plurality of documents into a plurality of clusters in accordance with semantic similarity; and

a cluster merging unit which-for evaluatinges the relation among the plurality of clusters created by said clustering unit on the basis of the documents included in the respective clusters and then combines two or more clusters having a degree of relation equal to or higher than a predetermined value;

wherein said cluster merging unit defines said degree of relation between multiple clusters under consideration as the number of distinct files common to all of said clusters under consideration multiplied by a predefined multiplication factor divided by a total sum of all the files in said clusters under consideration.

11. (Currently Amended) A document categorizing apparatus for categorizing a plurality of documents into a plurality of clusters according to semantic similarity, said apparatus comprising:

a clustering unit for categorizing a plurality of documents into a plurality of clusters in accordance with semantic similarity,

a cluster merging unit which for evaluatinges the relation among the plurality of clusters created by said clustering unit on the basis of the documents included in the respective clusters and then combines two or more clusters having a degree of relation equal to or higher than a first predetermined value;

~~a cluster merging process information generator for generating cluster merging process information representing which clusters have been merged together and also representing the degrees of relation among the merged clusters wherein said cluster merging process information is to be displayed when final clusters obtained via said cluster merging process performed by said cluster merging unit are displayed so that a user can see in what manner said cluster merging process has been performed to obtain said final cluster; and~~

~~categorization result outputting means for outputting said cluster merging process information such that said cluster merging process information is included in the categorization result to be presented to said user.~~

an output display for displaying the cluster names of respective clusters merged together such that when said degree of relation among said clusters is higher than a second predetermined value higher than said first predetermined value, said cluster names are displayed in a first listing format, and when said degree of relation among said clusters is lower than said second predetermined value and higher than said first predetermined value, said cluster names are displayed in a second listing format.

12. (Currently Amended) A storage medium on which a document categorizing program for categorizing a plurality of documents into a plurality of clusters according to semantic similarity is stored, said document categorizing program comprising:

a clustering step for categorizing a plurality of documents into a plurality of clusters in accordance with semantic similarity, and

a cluster merging step in which the degrees of relation among clusters of said plurality of clusters obtained in said clustering step are evaluated on the basis of documents included in the respective clusters, and two or more clusters having a degree of relation equal to or higher than a predetermined value are combined together;

wherein said cluster merging step defines said degree of relation between multiple clusters under consideration as the number of distinct files common to all of said clusters under consideration multiplied by a predefined multiplication factor divided by a total sum of all the files in said clusters under consideration.

13. (Currently Amended) A storage medium on which a document categorizing program for categorizing a plurality of documents into a plurality of clusters according to semantic similarity is stored, said document categorizing program comprising:

a clustering step for categorizing a plurality of documents into a plurality of clusters in accordance with semantic similarity;

a cluster merging step in which the degrees of relation among clusters of said plurality of clusters obtained in said clustering step are evaluated on the basis of documents included in the respective clusters, and two or more clusters having a degree of relation equal to or higher than a first predetermined value are combined together; and

a cluster merging process information generating step for generating cluster merging process information representing which clusters have been merged together and also representing the degrees of relation among the merged clusters wherein said cluster merging process information is to be displayed when final clusters obtained via said cluster merging process performed by said cluster merging step are displayed so that a user can see in what manner said cluster merging process has been performed to obtain said final cluster; and

a step for outputting said cluster-merging-process information such that said cluster merging process information is included in the categorization result to be presented to said user, wherein the cluster names of respective clusters merged together are displayed such that when said degree of relation among said clusters is higher than a second predetermined value higher than said first predetermined value, said cluster names are displayed in a first listing format, and when said degree of relation among said clusters is lower than said second predetermined value and higher than said first predetermined value, said cluster names are displayed in a second listing format.

14. (New) The document categorizing method of Claim 1, wherein said multiplication factor and said number of clusters under consideration is two.

15. (New) The document categorizing method of Claim 7, wherein said first listing format is an AND listing format and said second listing format is an OR listing format.

16. (New) The document categorizing apparatus of claim 10, wherein said multiplication factor is equal to the number of clusters under consideration.

17. (New) The document categorizing apparatus of Claim 10, wherein said multiplication factor and said number of clusters under consideration is two.

18. (New) The storage medium of Claim 12, wherein said multiplication factor is equal to the number of clusters under consideration.

19. (New) The storage medium of Claim 12, wherein said multiplication factor and said number of clusters under consideration is two.

20. (New) The document categorizing apparatus of claim 11, wherein said first listing format is an AND listing format and said second listing format is an OR listing format.

21. (New) The document categorizing apparatus of claim 11, wherein when said cluster names are displayed in said first listing format, said cluster names of the respective clusters are displayed successively in a single horizontal line or are displayed successively in different lines, and when said cluster names are displayed in said second listing format, a delimiter is inserted between adjacent cluster names of the respective clusters.

22. (New) The storage medium of claim 13, wherein said first listing format is an AND listing format and said second listing format is an OR listing format.

23. (New) The storage medium of claim 13, wherein when said cluster names are displayed in said first listing format, said cluster names of the respective clusters are displayed successively in a single horizontal line or are displayed

successively in different lines, and when said cluster names are displayed in said second listing format, a delimiter is inserted between adjacent cluster names of the respective clusters.